UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/660,326	09/10/2003	Dinei A. Florencio	MCS-033-03 (304925.01)	5957

27662 7590 01/04/2008

MICROSOFT CORPORATION
C/O LYON & HARR, LLP
300 ESPLANADE DRIVE
SUITE 800
OXNARD, CA 93036

EXAMINER	
LERNER, MARTIN	
ART UNIT	**PAPER NUMBER**
2626	

MAIL DATE	DELIVERY MODE
01/04/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/660,326	FLORENCIO ET AL.	
	Examiner	Art Unit	
	Martin Lerner	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *10 December 2007*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1 to 36* is/are pending in the application.

4a) Of the above claim(s) *15 to 36* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1 to 4 and 8 to 14* is/are rejected.

7)☒ Claim(s) *5 to 7* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *08 September 2007* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Election/Restrictions

Applicants' election of Group I, Claims 1 to 14, in the reply filed on 10 December 2007 is acknowledged. Because Applicants did not distinctly and specifically point out the supposed errors in the restriction requirement, the election has been treated as an election without traverse (MPEP § 818.03(a)).

Claims 15 to 36 are withdrawn from further consideration pursuant to 37 CFR 1.142(b) as being drawn to a nonelected invention, there being no allowable generic or linking claim. Election was made **without** traverse in the reply filed on 10 December 2007.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1 to 4 and 10 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Li et al.* in view of *Clemm*.

Concerning independent claim 1, *Li et al.* discloses a system for encoding speech, comprising:

"analyzing sequential segments of at least one digital audio signal to determine segment type as one of speech type segments, non-speech type segments, and unknown type segments" – a comparison of a current frame's ("sequential segments") full-band energy to a reference level is made; if the current frame's energy equals or exceeds the reference level, then a G.729 Annex B VAD (voice activity detector) sets an output to indicate the detected presence of voice activity in the current frame; if the current frame's energy is less than the reference level, a G.729 Annex B VAD sets its output to zero to indicate the non-detection of voice activity in the current frame (column 9, lines 48 to 63: Figure 3: Steps 23 to 27);

"encoding each speech segment as one or more signal frames using a speech segment-specific encoder" – if VAD 1 detects voice activity, a G.729 speech encoder 3 is invoked to encode the digital representation of the detected voice signal (column 1, line 63 to column 2, line 2: Figure 1); a G.729 encoder is "a speech segment-specific encoder";

"encoding each non-speech frame as one or more signal frames using a non-speech segment-specific encoder" – however, if VAD 1 does not detect voice activity, a Discontinuous Transmission/Comfort Noise Generator (noise) encoder 2 is used to code the digital representation of the detected background noise signal (column 1, line 63 to column 2, line 2: Figure 1); a Discontinuous Transmission/Comfort Noise Generator (noise) encoder 2 is "a non-speech segment-specific encoder".

Concerning independent claim 1, *Li et al.* further discloses that G.729 Annex B performs a multi-boundary initial G.729 Annex B decision to refine an initial decision to

reflect the long-term stationary nature of the voice signal. After the initial VAD decision has been smoothed, a final decision is formed. (Column 2, Lines 30 to 38; Column 10, Lines 24 to 37) Thus, *Li et al.* discloses that whether or not a current frame's full-band energy exceeds the reference value is only a first step in determining voice activity, so that there may be frames, through further refinement of the decision, that are equivalent to "unknown type segments". Figure 3 shows that after VAD voice detection is set to 1 or 0, Figure 4 continues with a multi-boundary initial VAD decision to make a smoothed final VAD decision due to background noise, which may change the value of the VAD decision from 0 to 1. If the running averages of the background noise characteristics and supplemental VAD algorithms have diverged, then the values for these characteristics generated by the supplemental VAD algorithm are substituted for the respective values of these characteristics generated by the G.729 Annex B algorithm. (Column 12, Lines 5 to 12: Figure 4: Steps 30 and 41) Subsequently, either speech encoder 3 or Discontinuous Transmission/Comfort Noise Generator (noise) encoder 2 is used to code the digital representation of the voice signal or background noise signal according to the refinement of the final decision. (Column 1, Line 63 to Column 2, Line 2: Figure 1)

Concerning independent claim 1, the only elements omitted by *Li et al.* are the steps of "buffering each sequential unknown type segment in a segment buffer until analysis of a subsequent segment identifies the subsequent segment type as any of a speech segment and silence segment" and "encoding the buffered segments and the subsequent segments as one or more signal frames using the segment-specific

encoder corresponding to the type of the subsequent segment." Generally, it is known that speech encoding involves buffering during processing, implicitly, but buffering is not expressly disclosed by *Li et al.*, and *Li et al.* makes a final decision based upon a running average of energy in previous segments instead of subsequent segments. However, *Clemm* teaches VAD-directed silence suppression, where a voice signal is received in a buffer during a delay between a start of voice activity and the detection of voice activity. (Column 1, Lines 28 to 58: Figure 1) An objective of buffering voice signals is to ensure that no voice activity is lost during the period of time necessary to turn off silence suppression. (Column 1, Lines 49 to 54) *Clemm* relies upon information from subsequent segments to decide whether a segment ("each sequential unknown type segment") is the start of voice activity ("a speech segment") or a continuation of silence ("a silence segment"). (Column 2, Line 59 to Column 3, Line 30: Figure 2) Once an ambiguous segment, *i.e.* segment bd in Figure 2 of *Clemm*, is declared to be a speech segment, then *Li et al.* would code the segment as speech with a G.729 speech encoder ("encoding the buffered segments and the subsequent segment as one or more signal frames using the segment-specific encoder corresponding to the type of the subsequent segment"). It would have been obvious to one having ordinary skill in the art to buffer segments and encode based upon subsequent segments as suggested by *Clemm* in a method for encoding voice activity by G.729 Annex B of "unknown type segments" of *Li et al.* for a purpose of ensuring that no voice signals are lost during a period of time to determine whether a speech signal has voice activity or no voice activity.

Concerning claim 2, *Li et al.* discloses detection of voice activity and non-voice activity for background noise (column 1, line 63 to column 2, line 2: Figure 1); initially, a current frame's energy is compared to a reference level to determine whether voice activity is detected (column 9, lines 48 to 63); thus, an initial decision reflects whether the current frame is speech or silence.

Concerning claim 3, *Clemm* teaches that the speed of playback may be increased to 150% speed playback when the buffer is full, according to the depletion level of the buffer (column 3, lines 30 to 42: Figures 3A and 3B); increasing the speed of playback before transmission is equivalent to "a burst transmission at a higher rate than a current sampling rate of the audio signal".

Concerning claim 4, *Clemm* teaches a depletion device flushes the buffer in an accelerated manner when the VAD function is released (column 4, lines 38 to 39).

Concerning claim 10, *Li et al.* discloses a decoder for receiving voice and noise encoded signals (column 2, lines 7 to 14: Figure 1); *Clemm* teaches "a burst transmission" as a speed of playback may be increased to 150% speed playback when the buffer is full, according to the depletion level of the buffer before transmission by transmission device 450 (column 3, lines 30 to 42: Figures 1, 3A, 3B, and 4); implicitly, *Li et al.* operates at a fixed frame rate.

Claims 8 to 9 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Li et al.* in view of *Clemm* as applied to claim 1 above, and further in view of *Lakaniemi et al.*

Li et al. discloses that an initial VAD decision is made using multi-boundary decision regions, but does not expressly disclose identifying an onset point of speech, where an onset point can be identified and encoded as non-speech segments or as speech segments. However, it is known to classify speech frames in a variety of ways, including as onset frames. *Lakaniemi et al.* teaches classifying speech frames into frame types, with frames having lower priority, such as non-speech frames, being selected for control message data, and frames having higher priority frame types, such as onset and transient frames, being avoided for selection due to the higher subjective contribution to speech quality. (Abstract) (*Clemm* teaches the features of buffering, temporally compressing, and discarding frames, as noted above.) It would have been obvious to one having ordinary skill in the art to identify an actual onset point of speech in a current segment as taught by *Lakaniemi et al.* in a method for encoding voice activity by G.729 Annex B of *Li et al.* for a purpose of avoiding selecting frames having a higher subjective contribution to speech quality for control message data.

Claims 11 to 14 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Li et al.* in view of *Clemm* as applied to claims 1, 3, and 10 above, and further in view of *Ramjee et al. ("Adaptive Playout Mechanisms for Packetized Audio Applications in Wide-Area Networks")*.

Li et al. omits a decoder that uses extra samples contained in a burst transmission to populate a jitter buffer for an adaptive playout scheme, where at least some of the received data is compressed to reduce average signal delay. However,

Ramjee et al. teaches an adaptive playout mechanism, where received audio packets are buffered, and their playout delayed at the destination host in order to compensate for variable network delays. (Abstract) *Ramjee et al.* discloses a delay jitter (Page 2, Left Column) for a buffer having a maximum size (Page 1, Right Column), which is equivalent to a "jitter buffer". The algorithm is applied to talkspurts, which are equivalent to "burst transmission". It would have been obvious to one having ordinary skill in the art to employ the adaptive playout mechanism with a jitter buffer of *Ramjee et al.* in a method for encoding voice activity by G.729 Annex B of *Li et al.* for a purpose of compensating for variable network delays.

Allowable Subject Matter

Claims 5 to 7 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Response to Arguments

Applicants' arguments filed 08 September 2007 have been fully considered but they are not persuasive.

Applicants argue that *Li et al.* makes a final VAD decision of whether a segment is speech or non-speech based on the relationship between the detected energy of neighboring past frames by taking running averages of past segments. Applicants say that their system for encoding an audio signal buffers only unknown segments.

Moreover, Applicants maintain that neither *Li et al.* nor *Clemm* buffers unknown segments until an analysis of a subsequent frame identifies the subsequent segment type as either a speech segment and a silence segment, at which time the buffered segments are encoded as either speech or non-speech based on the type of the subsequent frame. Applicants' main argument, then, is that *Li et al.* makes a decision as to whether to classify the segment as speech or non-speech based on previous frames, while their system for encoding an audio signal classifies the segment as speech or non-speech based on subsequent frames. Applicants' argument is being considered after careful review of the references, but is not persuasive.

Firstly, Applicants' implication that their system for encoding only provides for buffering of unknown segments is neither claimed nor would one having ordinary skill in the art find it to be necessarily technically accurate. It is maintained that any system for encoding speech would buffer not just the unknown segments but any known speech and known non-speech segments, too. Speech and non-speech segments must be buffered, too, because coding of speech and non-speech segments requires processing to produce a code for the speech and non-speech segments. Any processing of the speech and non-speech segments would require these segments to be placed in some temporary storage location while they are being processed for coding. Thus, it would not be correct to say that only the unknown segments are buffered. Moreover, the limitation that only the unknown segments are buffered, as implied by the argument of Applicants, is not any express element of the independent claim. It follows that any

implication that the prior art fails to disclose buffering only the unknown segments is not persuasive.

Secondly, *Clemm* teaches the limitation of buffering unknown segments until analysis of a subsequent segment identifies a subsequent segment as any of a speech segment and a silent segment, and then declaring that a current buffered unknown segment is of the same type as a subsequent segment. Upon care review of that reference, *Clemm* does more than simply suggest buffering for processing speech in a VAD. *Clemm* discloses that a current frame and a subsequent frame are buffered. If a current frame is buffered after silence, it may not be clear whether the current frame is the start of voice activity or more silence. So, *Clemm* looks at subsequent frames – which is known in the art as "lookahead" – stored in the buffer for detection of voice activity. Then, if there is a delay between the point at which voice activity actually started in the current frame and when the voice activity was detected in a subsequent frame, the effect of clipping the start of speech is greatly reduced. (Column 2, Lines 28 to 46)

This is best understood by looking at segment bd in Figure 2 of *Clemm*. Segment bd is the actual start of voice activity following a silence segment between 241 and 242. However, because there may be a delay in recognizing when speech actually begins, silence suppression may ordinarily end at 243, thereby clipping the beginning of speech at segment bd between 242 and 243. Still, by storing a current segment bd in a buffer, and looking at all the subsequent segments of speech between 242 and 244, it becomes apparent that speech actually began at 242. (Column 2, Line 59 to Column 3,

Line 30: Figure 2) Because an initially unknown segment bd was stored in the buffer until subsequent segments were definitively declared to be speech, it is feasible, at the expense of a brief delay in transmission, to eliminate clipping at the beginning of speech, and retroactively declare a current segment, bd, to be speech based upon analysis of subsequent segments. Thus, *Clemm* teaches Applicants' limitation of "buffering each sequential unknown type segment in a segment buffer until analysis of a subsequent segment identifies the subsequent segment as any of a speech segment and a silence segment", which is the limitation that Applicants argue is omitted by *Li et al.*

Admittedly, Applicants are correct in saying that *Li et al.* detects voice activity of ambiguous segments based upon a running average of energy in previous segments instead of subsequent segments. However, *Clemm* makes up for the omission of *Li et al.* by providing a "lookahead" classification of segments as speech or silence based upon analysis of subsequent segments instead of analysis of previous segments.

Moreover, any argument that *Li et al.* teaches away by classifying ambiguous segments as speech or silence based on previous segments instead of subsequent segments would not be persuasive because that would simply be considering the two references individually without consideration of what is taught by *Clemm* as well as *Li et al.* Indeed, *Li et al.* discloses that some segments are clearly speech and some segments are clearly silence, and that speech segments are encoded with a first encoding method (a G.729 speech encoder) and that silence segments are encoded with a second encoding method (a Discontinuous Transmission/Comfort Noise

Generator (noise) encoder). (Column 1, Line 63 to Column 2, Line 2: Figure 1) That is standard for Recommendation G.729 Annex B. *Li et al.* then goes on to describe how multi-boundary decisions are made for segments that are ambiguous between speech and non-speech. Once a final decision is made that the segment is speech or non-speech, though, *Li et al.* encodes that segment with either the first encoding method or the second encoding method pursuant to that final decision. Thus, *Li et al.* suggests the limitation of "encoding the buffered segments . . . as one or more signal frames using the segment-specific encoder corresponding to the type of" one of the frame types.

Li et al. makes the final decision based on energy information from previous frames instead of subsequent frames, but *Clemm* provides the teaching to buffer a current frame until it becomes clear from a subsequent frame what frame type a current frame was. Moreover, *Clemm* is concerned with solving the same problem in the same way as Applicants. Any delay introduced by buffering to preserve the beginning of speech is compensated by compressing at least a portion of the segment contained in the buffer. Thus, it would have been obvious to one having ordinary skill in the art to encode an unknown segment based upon a segment type of a subsequent segment as taught by *Clemm* instead of encoding an unknown segment based upon a segment type of a previous segment as taught by *Li et al.*

Therefore, the rejections of claims 1 to 4 and 10 under 35 U.S.C. §103(a) as being unpatentable over *Li et al.* in view of *Clemm*, of claims 8 to 9 under 35 U.S.C. §103(a) as being unpatentable over *Li et al.* in view of *Clemm*, and further in view of *Lakaniemi et al.*, and of claims 11 to 14 under 35 U.S.C. §103(a) as being

unpatentable over *Li et al.* in view of *Clemm*, and further in view of *Ramjee et al.*, are proper.

Conclusion

The prior art made of record and not relied upon is considered pertinent to Applicants' disclosure.

Lee et al. discloses related art.

THIS ACTION IS MADE FINAL. Applicants are reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

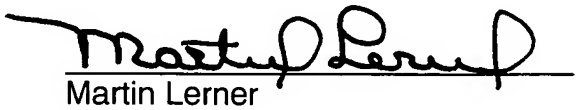
A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (571) 272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

ML
12/21/07



Martin Lerner
Examiner
Group Art Unit 2626